Nearly Unstable Processes: A Prediction Perspective

Ngai Hang Chan

The Chinese University of Hong Kong

Ching-Kang Ing

National Tsing Hua University

Rongmao Zhang

Zhejiang University

Supplementary Material

Section S1 contains omitted proofs of the theorems in Sections 2 and 3.1, whereas Section S2 provides practical guidelines for choosing finite sample approximations from those derived in the near unit-root and the general near unit-root models.

S1 Proofs of the Theorems in Sections 2 and 3.1

Before proceeding with the proofs, we would like to first point out the key difference between our asymptotic framework and those in Chan and Wei (1987), Phillips (1987) and Phillips and Magdalinos (2007). Let $\rho_n = 1 - b/n^{\beta}$, with $0 < \beta \leq 1$ and $0 < b < \infty$, and $y_t^* = y_t - n^{\beta}\mu/b$. Then, model (1.1) can be expressed as

$$y_t^* = \rho_n y_{t-1}^* + \eta_t, \tag{S1.1}$$

which is an AR(1) model driven by a zero-mean white-noise process $\{\eta_t\}$. While the asymptotic behavior of the LSE under (S1.1) has been extensively studied by the aforementioned authors, their results rely heavily on the initial condition $y_0^* = O_p(1)$, which is obviously violated by our initial condition $y_0 = 0$, leading to

$$y_0^* = -n^\beta \mu/b. \tag{S1.2}$$

With the initial condition like (S1.2), most existing results established for the LSE under model (S1.1) are no longer applicable. As shown for the rest of this section, our asymptotic analysis is similar to that adopted by Ing and Yang (2014). However, substantial efforts are needed to deal with the critical behavior of the EV and LS predictors exhibited in the near unit-root region.

Proof of Theorem 2.1. We first prove (2.6). By $\exp(-b)\sum_{j=1}^t \varepsilon_j \le y_t = \sum_{j=0}^{t-1} \rho_n^j \varepsilon_{t-j} \le \sum_{j=1}^t \varepsilon_j$ and (19) of Ing and Yang (2014), which shows that for any $q > 0$,

$$\mathrm{E}\left\{n^{1+1/\alpha} \min_{2\le t\le n}\left(\varepsilon_t / \sum_{j=1}^{t-1} \varepsilon_j\right)\right\}^q < \infty, \tag{S1.1}$$

the desired conclusion (2.6) follows. The proof of (2.7) is similar to that of (7) of Ing and Yang (2014). The details are omitted. Finally, (2.8) follows directly from (2.6) and (2.7). □

Proof of Theorem 3.1. Equation (3.1) can be shown by an argument similar to that used to prove (2.6). We thus skip the details. For (3.2), it suffices to show that

$$n^{1/\alpha+\beta} \min_{\nu_n \le i \le n} (\varepsilon_i/y_{i-1}) - n^{1/\alpha+\beta} \min_{2 \le i \le n} (\varepsilon_i/y_{i-1}) = o_p(1) \tag{S1.2}$$

and for any $t > 0$,

$$\lim_{n\to\infty} P\{(c/\alpha)^{1/\alpha}(\mu/b) n^{1/\alpha+\beta} \min_{\nu_n \le i \le n} (\varepsilon_i/y_{i-1}) > t\} = \exp\{-t^\alpha\}, \tag{S1.3}$$

where $v_n \asymp n^\theta$ for some $\beta < \theta < 1$.

To show (S1.2), note first that

$$n^{1/\alpha+\beta} \min_{v_n \le i \le n} (\varepsilon_i/y_{i-1}) - n^{1/\alpha+\beta} \min_{2 \le i \le n} (\varepsilon_i/y_{i-1}) \le n^{1/\alpha+\beta} (\min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1}) I_{A_n},$$

where $A_n = \{\min_{\nu_n \le i \le n} \varepsilon_i/y_{i-1} > \min_{2 \le i \le \nu_n} \varepsilon_i/y_{i-1}\}$. Let $q_n = s_n^{1/2} n^\beta/\nu_n^\beta$ in which s_n satisfies $s_n \nu_n^{1/\alpha}/n^{1/\alpha} = o(1)$ and $s_n \to \infty$. Then, by (3.1), (1.2), the weak law of large number and Chebyshev's inequality, one has for any

$\epsilon > 0$,

$$
\begin{aligned}
& P(n^{1/\alpha+\beta}(\min_{\nu_n \le i \le n} \varepsilon_i / y_{i-1}) I_{A_n} > \epsilon) \le P(A_n) \\
\le \;& P(\min_{\nu_n \le i \le n} \varepsilon_i / y_{i-1} > s_n^{-1/2} q_n^{-1} \nu_n^{-\beta-(1/\alpha)}) + P(\max_{2 \le i \le \nu_n} y_{i-1} \ge q_n \nu_n^{\beta}) \\
+ \;& P(\min_{2 \le i \le \nu_n} \varepsilon_i < s_n^{-1/2} \nu_n^{-1/\alpha}) \\
= \;& O\left(\frac{s_n^{1/2} q_n \nu_n^{1/\alpha+\beta}}{n^{1/\alpha+\beta}}\right) + o(1) + 1 - \left(1 - \frac{C}{\nu_n s_n^{\alpha/2}}\right)^{\nu_n} = o(1),
\end{aligned}
$$

where C is some positive constant independent of n. Thus, (S1.2) is proved.

To show (S1.3), one can use $\mathrm{E}(\varepsilon_1^{q_1}) < \infty$, for some $q_1 > 2/\beta$, and Lemma 2 of Wei (1987) to obtain

$$
\max_{\nu_n \le i \le n} |y_{i-1} - \mathrm{E} y_{i-1}| = O_p(n^{1/q_1+\beta/2}). \tag{S1.4}
$$

In addition, there exists $c_1 > 0$ such that for all $\nu_n \le i \le n$,

$$
\mathrm{E} y_{i-1} = \mu n^{\beta}(1-\rho^i)/b \ge c_1 n^{\beta}, \tag{S1.5}
$$

By (S1.4), (S1.5), (1.2) and $1/q_1 + \beta/2 < \beta$, it holds that

$$
\begin{aligned}
& \lim_{n\to\infty} P\{(c/\alpha)^{1/\alpha}(\mu/b) n^{1/\alpha+\beta} \min_{\nu_n \le i \le n} (\varepsilon_i / y_{i-1}) > t\} \\
& = \lim_{n\to\infty} P\left\{(\mu/b)(c/\alpha)^{1/\alpha} n^{1/\alpha+\beta} \min_{\nu_n \le i \le n} \left(\frac{\varepsilon_i}{\mu n^{\beta}(1-\rho^i)/b}\right) > t\right\} = \exp\{-t^{\alpha}\},
\end{aligned}
$$

which completes the proof of (S1.3). Finally, (3.3) is a immediate consequence of (3.1) and (3.2). $\square$

Proofs of Theorems 2.2 and 3.2. Define

$$a_{n,i} = \left(\frac{1-\rho_n^i}{1-\rho_n} - \frac{1}{(n-1)} \sum_{j=1}^{n-1} \frac{1-\rho_n^{j-1}}{1-\rho_n} \right) \mu,$$

and $\xi_{n,i} = \sum_{j=0}^{i-1} \rho^j \eta_{i-j} - (n-1)^{-1} \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \rho^j \eta_{i-j}$. Then

$$\begin{aligned} \tilde{\rho}_n - \rho_n &= \sum_{i=2}^{n} (y_{i-1} - \bar{y}) \eta_i / \sum_{i=2}^{n} (y_{i-1} - \bar{y})^2 \\ &= \sum_{i=2}^{n} (a_{n,i-1} + \xi_{n,i-1}) \eta_i / \sum_{i=2}^{n} (a_{n,i-1} + \xi_{n,i-1})^2. \end{aligned} \tag{S1.6}$$

Straightforward calculations yield for $0 < \beta \leq 1$,

$$\lim_{n\to\infty} \frac{1}{n^{3\beta}} \sum_{i=2}^{n} a_{n,i-1}^2 = \frac{\mu^2}{2b^3} I(0<\beta<1) + \mu^2 (I_1(b) - I_2(b)) I(\beta = 1), \tag{S1.7}$$

and

$$\sum_{i=2}^{n} \mathrm{E}\xi_{n,i-1}^2 = \sigma^2 (2b)^{-1} n^{1+\beta} (1 - \exp(-bn^{1-\beta}))(1 + o(1)). \tag{S1.8}$$

In addition, for $0 < \beta < 1$, we have

$$\frac{1}{n^{1+\beta}} \sum_{i=2}^{n} \xi_{n,i-1}^2 - \frac{1}{n^{1+\beta}} \sum_{i=2}^{n} (\sum_{j=0}^{i-1} \rho^j \eta_{i-j})^2 = o_p(1), \tag{S1.9}$$

and

$$\frac{1}{n^{1+\beta}} \sum_{i=2}^{n} (\sum_{j=0}^{i-1} \rho^j \eta_{i-j})^2 \xrightarrow{p} \frac{\sigma^2}{2b}. \tag{S1.10}$$

Moreover, for $\beta = 1/2$,

$$\begin{aligned} &\frac{1}{n^{3/2}} \sum_{i=2}^{n} (a_{n,i-1} + \xi_{n,i-1})^2 = \frac{1}{n^{3/2}} \sum_{i=2}^{n} (a_{n,i-1}^2 + \xi_{n,i-1}^2) + o_p(1) \\ &= \frac{\sigma^2}{2b} + \frac{\mu^2}{2b^3} + o_p(1). \end{aligned} \tag{S1.11}$$

By (S1.6)–(S1.11) and the martingale central limit theorem (see, e.g., Theorem 3.2 of Phillips and Magdalinos (2007)), (2.10) and (3.4) follow.

Set $k_n = n^3$ for $\beta = 1$. By (2.11), $\mathrm{E}\varepsilon_1^s < \infty$ for some $s > 10$, and an argument similar to that used to prove Lemma 2 of Yu, Lin and Cheng (2012), we have $\mathrm{E}|\sqrt{k_n}(\tilde{\rho}_n - \rho_n)|^\gamma < \infty$ for some $\gamma > 2$, and hence $\{k_n(\tilde{\rho}_n - \rho_n)^2\}$ is uniformly integrable. This together with (2.10) (resp. (3.4)) yields (2.12) (resp. (35)). □

Proofs of Theorems 2.3 and 3.3. It follows from (2.4) and $\mathrm{E}[(1-\rho_n)\bar{y} - \mu + z_n]^2 = \mathrm{E}[(n-1)^{-1}\sum_{j=2}^n \eta_j]^2 = \sigma^2/(n-1)$ that

$$\begin{aligned} \mathrm{MSPE}_A - \sigma^2 &= \sigma^2/(n-1) + \mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})\}^2 \\ &\quad + 2\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})[((1-\rho_n)\bar{y} - \mu) + z_n]\} \end{aligned} \tag{S1.12}$$

To deal with the second term on the right-hand side of (S1.12), we express $y_n - \bar{y}$ as

$$\begin{aligned} y_n - \bar{y} &= \left(\sum_{j=0}^{n-1} \rho^j - \frac{1}{n-1}\sum_{i=1}^{n-1}\sum_{j=0}^{i-1}\rho^j\right)\mu + \left(\sum_{j=0}^{n-1}\rho^j\eta_{n-j} - \frac{1}{n-1}\sum_{i=1}^{n-1}\sum_{j=0}^{i-1}\rho^j\eta_{i-j}\right) \\ &:= X_{1,n} + X_{2,n}. \end{aligned} \tag{S1.13}$$

Some algebraic manipulations give

$$\lim_{n\to\infty} \frac{X_{1,n}^2}{n^{4\beta-2}} = \frac{\mu^2}{b^4} I(0 < \beta < 1) + \mu^2 L_3(b) I(\beta = 1) \tag{S1.14}$$

and

$$\mathrm{E}X_{2,n}^2 = \frac{\sigma^2}{2b} n^{\beta}(1+o(1))I(0<\beta<1) + O(n)I(\beta=1). \tag{S1.15}$$

Combining (S1.13)–(S1.15) yields for $2/3 < \beta \le 1$,

$$\frac{(y_n-\bar{y})^2}{n^{4\beta-2}} \xrightarrow{p} \frac{\mu^2}{b^4} I(2/3<\beta<1) + \mu^2 L_3(b) I(\beta=1), \tag{S1.16}$$

and for $0 < \beta \le 2/3$,

$$\lim_{n\to\infty} n^{-\beta}\mathrm{E}(y_n-\bar{y})^2 = \frac{\sigma^2}{2b} I(0<\beta<2/3) + \frac{\mu^2}{b^4} I(\beta = 2/3). \tag{S1.17}$$

By the moment conditions on ε_1 and a straightforward calculation, it follows that for $0<\beta\le 2/3$ there exists $2/3<\zeta<1$ for which

$$n^{-\beta/2}(y_n-\bar{y}) = n^{-\beta/2}(X_{1,n} + \sum_{j=0}^{n^{\zeta}-1} \rho^j \eta_{n-j}) + r_{1,n},$$

where $r_{1,n}$ satisfies $\mathrm{E}|r_{1,n}|^{q_1} = o(1)$ for some $q_1>2$, and

$$\lim_{n\to\infty} \mathrm{E}[n^{-\beta}(X_{1,n} + \sum_{j=0}^{n^{\zeta}-1} \rho^j \eta_{n-j})^2] = \lim_{n\to\infty} \mathrm{E}[n^{-\beta}(y_n-\bar{y})^2].$$

In addition, by (3.1) and an argument similar to that used to prove (S1.2), we obtain

$$n^{\beta+1/\alpha}(\hat{\rho}_n-\rho_n) = n^{\beta+1/\alpha}\Big(\min_{2\le i\le n-n^{\zeta}} \frac{\varepsilon_i}{y_{i-1}}\Big) + r_{2,n},$$

where $r_{2,n}$ satisfies $\mathrm{E}|r_{2,n}|^q = o(1)$ for any $q>0$, and

$$\lim_{n\to\infty} \mathrm{E}\Big[n^{\beta+1/\alpha}\Big(\min_{2\le i\le n-n^{\zeta}} \frac{\varepsilon_i}{y_{i-1}}\Big)\Big]^2 = \lim_{n\to\infty} \mathrm{E}[n^{\beta+1/\alpha}(\hat{\rho}_n-\rho_n)]^2.$$

These facts and the independence between $n^{-\beta/2}\sum_{j=0}^{n^\zeta-1}\rho^j\eta_{n-j}$ and $n^{\beta+1/\alpha}(\min_{2\le i\le n-n^\zeta}\varepsilon_i/y_{i-1})$ yield for $0<\beta\le 2/3$

$$\begin{aligned}&\lim_{n\to\infty}\mathrm{E}\{n^{-\beta}(y_n-\bar{y})^2[n^{\beta+1/\alpha}(\hat{\rho}_n-\rho_n)]^2\}\\&=\lim_{n\to\infty}\mathrm{E}\{n^{-\beta}(y_n-\bar{y})^2\}\lim_{n\to\infty}\mathrm{E}\{[n^{\beta+1/\alpha}(\hat{\rho}_n-\rho_n)]^2\}.\end{aligned}\tag{S1.18}$$

Now, by (S1.16)–(S1.18), (2.6), (2.8), (3.1), (3.3) and the moment conditions imposed on ε_1, it holds that for $2/3<\beta\le 1$,

$$\begin{aligned}&\lim_{n\to\infty}\mathrm{E}[n^{-\beta+1/\alpha+1}(\hat{\rho}_n-\rho_n)(y_n-\bar{y})]^2\\&=\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{2/\alpha}b^{-2}I(2/3<\beta<1)\\&+\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{cM'_{\alpha,b}}\Big)^{2/\alpha}L_3(b)I(\beta=1),\end{aligned}\tag{S1.19}$$

and for $0<\beta\le 2/3$,

$$\begin{aligned}&\lim_{n\to\infty}\mathrm{E}[n^{1/\alpha+\beta/2}(\hat{\rho}_n-\rho_n)(y_n-\bar{y})]^2\\&=\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{2/\alpha}\left\{\frac{\sigma^2 b}{2\mu^2}I(0<\beta\le 2/3)+\frac{1}{b^2}I(\beta=2/3)\right\}.\end{aligned}\tag{S1.20}$$

To deal with the third term on the right-hand side of (S1.12), we obtain from an argument similar to that used to prove (2.9) in the supplementary document for Ing and Yang (2014) that for $2/3<\beta\le 1$,

$$\begin{aligned}&\mathrm{E}\{(\hat{\rho}_n-\rho_n)(y_n-\bar{y})[((1-\rho_n)\bar{y}-\mu)+z_n]\}\\&=o(\max\{n^{-1},n^{2\beta-2-2/\alpha}\}),\end{aligned}\tag{S1.21}$$

and for $0 < \beta \leq 2/3$,

$$\begin{aligned} &\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})[((1-\rho_n)\bar{y} - \mu) + z_n]\} \\ &= o(\max\{n^{-1}, n^{-\beta-2/\alpha}\}). \end{aligned} \tag{S1.22}$$

Consequently, the desired conclusions (2.13) and (3.6)–(3.8) are ensured by (S1.12), and (S1.19)–(S1.22). □

Proofs of Theorems 2.4 and 3.4. It follows from (2.5) that

$$\begin{aligned} &\mathrm{E}(y_{n+1} - \tilde{y}_{n+1})^2 - \sigma^2 = \frac{\sigma^2}{n-1} + \mathrm{E}\{(y_n - \bar{y})(\tilde{\rho}_n - \rho_n)\}^2 \\ &+ \frac{2}{n-1}\mathrm{E}\{(\sum_{i=2}^{n} \eta_i)(y_n - \bar{y})(\tilde{\rho}_n - \rho_n)\}. \end{aligned} \tag{S1.23}$$

By (S1.7)–(S1.11), Theorems 2.2 and 3.2, (2.11), $\mathrm{E}\varepsilon_1^s < \infty$ for some $s > 12$, and an argument similar to that used to prove Lemma 2 of Yu, Lin and Cheng (2012), we obtain, after some tedious algebraic manipulations,

$$\mathrm{E}\{(\sum_{i=2}^{n} \eta_i)(y_n - \bar{y})(\tilde{\rho}_n - \rho_n)\} = o(1), \tag{S1.24}$$

and

$$\lim_{n\to\infty} \frac{n}{\sigma^2}\mathrm{E}\{(y_n - \bar{y})(\tilde{\rho}_n - \rho_n)\}^2 = \begin{cases} 1 & 0 < \beta < 1/2\,, \\ \frac{b^2\sigma^2}{\mu^2 + b^2\sigma^2} & \beta = 1/2\,, \\ 0 & 1/2 < \beta < 1\,, \\ \frac{L_3(b)}{I_1(b) - I_2(b)} & \beta = 1. \end{cases} \tag{S1.25}$$

Consequently, Theorems 2.4 and 3.4 are guaranteed by (S1.23)–(S1.25).

S2 The implementation of finite sample approximations

S2.1 Rules of thumb developed from Tables 1–6

With the help of Tables 1–3 (Tables 4–6), we offer a simple rule for choosing a better approximation of $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ $(n(\mathrm{MSPE}_B - \sigma^2))$ from $R_A^{(2)}$ and $R_A^{(3)}$ ($R_B^{(2)}$ and $R_B^{(3)}$) when $100 \leq n \leq 1000$, $1 \leq n(1-\rho) = b \leq 140$ and $1.5 \leq \alpha \leq 4$. According to Tables 1–3, we first introduce Rule I for approximating $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$:

Rule I.

1. Choose $R_A^{(3)}$ if $1.5 \leq \alpha < 2$.

2. Choose $R_A^{(2)}$ if $2 \leq \alpha \leq 4$ and $1 \leq b \leq 5$.

3. Choose $R_A^{(3)}$ if $2 \leq \alpha \leq 4$ and $5 < b \leq 140$.

Although there are a few cases where Rule I leads to a $P_A^{(i)}$ (defined in Section 3.2) slightly smaller than $\max_{1\leq i\leq 3} P_A^{(i)}$, the rule has the advantage of easy implementation, which is practically appealing. In the same spirit, we propose using Rule II (according to Tables 4–6) for approximating $n(\mathrm{MSPE}_B - \sigma^2)$:

Rule II.

1. Choose $R_B^{(2)}$ if $1.5 \leq \alpha \leq 2$ and $1 \leq b \leq 12.5$.

2. Choose $R_B^{(3)}$ if $1.5 \leq \alpha \leq 2$ and $12.5 < b \leq 140$.

3. Choose $R_B^{(2)}$ if $2 < \alpha \leq 4$ and $1 \leq b \leq 25$.

4. choose $R_B^{(3)}$ if $2 < \alpha \leq 4$ and $25 < b \leq 140$.

Note that Rules I and II can be further refined by checking a more dense grid of n, b and α, which is not pursued here. In Section B.2, we provide reliable estimators, $\hat{\rho}_n^*$, $\hat{b}_n^*$, $\hat{\alpha}_n^*$, $\hat{c}_n^*$, $\hat{\mu}_n^*$ and $\hat{\sigma}_n^{2^*}$, of ρ, b, α, c, μ and σ^2. With these estimators, Rules I and II can be implemented in practice via replacing b, α, $R_A^{(i)}, i = 2, 3$ and $R_B^{(i)}, i = 2, 3$ therein by $\hat{b}_n^*$, $\hat{\alpha}_n^*$, $\hat{R}_A^{(i)}, i = 2, 3$ and $\hat{R}_B^{(i)}, i = 2, 3$, where

$$\hat{R}_A^{(2)} = \begin{cases} \Gamma\left(\frac{\hat{\alpha}_n^*+2}{\hat{\alpha}_n^*}\right)\left(\frac{\hat{\alpha}_n^*}{\hat{c}_n^*\hat{M}'_{\hat{\alpha}_n^*,\hat{b}_n^*}}\right)^{2/\hat{\alpha}_n^*} L_3(\hat{b}_n^*) + \hat{\sigma}_n^{*^2} I(\hat{\alpha}_n^* = 2) & \hat{\alpha}_n^* \geq 2\,, \\ \hat{\sigma}_n^{*^2} & \hat{\alpha}_n^* < 2, \end{cases} \tag{S2.26}$$

in which $\hat{M}'_{\hat{\alpha}_n^*,\hat{b}_n^*}$ is $M'_{\alpha,b}$ with α and b replaced by $\hat{\alpha}_n^*$ and $\hat{b}_n^*$, respectively,

$$\hat{R}_A^{(3)} = \begin{cases} \hat{R}_1^* + \hat{R}_2^* + \frac{\hat{\sigma}_n^{*^2}}{n^{1-2/\hat{\alpha}_n^*}} & \hat{\alpha}_n^* \geq 2\,, \\ n^{1-2/\hat{\alpha}_n^*}(\hat{R}_1^* + \hat{R}_2^*) + \hat{\sigma}_n^{*^2} & \hat{\alpha}_n^* < 2, \end{cases} \tag{S2.27}$$

with

$$\hat{R}_1^* = \Gamma((\hat{\alpha}_n^* + 2)/\hat{\alpha}_n^*)(\hat{\alpha}_n^*/\hat{c}_n^*)^{2/\hat{\alpha}_n^*}[\hat{\sigma}_n^{*2}(1-\hat{\rho}_n^*)]/2\hat{\mu}_n^{*2},$$

and

$$\hat{R}_2^* = \Gamma((\hat{\alpha}_n^* + 2)/\hat{\alpha}_n^*)(\hat{\alpha}_n^*/\hat{c}_n^*)^{2/\hat{\alpha}_n^*}[n(1-\hat{\rho}_n^*)]^{-2},$$

$$\hat{R}_B^{(2)} = \left\{1 + \frac{L_3(\hat{b}_n^*)}{I_1(\hat{b}_n^*) - I_2(\hat{b}_n^*)}\right\}\hat{\sigma}_n^{*2}, \tag{S2.28}$$

and

$$\hat{R}_B^{(3)} = \left\{1 + \frac{\hat{b}_n^{*2}\hat{\sigma}_n^{*2}}{n\hat{\mu}_n^{*2} + \hat{b}_n^{*2}\hat{\sigma}_n^{*2}}\right\}\hat{\sigma}_n^{*2}. \tag{S2.29}$$

Estimation of unknown parameters in Rules I and II

In this section, we address the problem of estimating the unknown parameters in Rules I and II. Suppose first that $\alpha > 2$ or $\alpha \le 2$ is known a priori. Then, according to Theorems 1 and 2 and Remark 3, it is reasonable to estimate ρ by

$$\hat{\rho}_n^* = \begin{cases} \hat{\rho}_n & \alpha \le 2\,, \\ \tilde{\rho}_n & \alpha > 2. \end{cases} \tag{S2.30}$$

By virtue of (S2.30), it is natural to estimate μ and σ^2 by $\hat{\mu}_n^* = n^{-1}\sum_{t=1}^{n-1}(y_{t+1} - \hat{\rho}_n^* y_t)$ and $\hat{\sigma}_n^{*2} = n^{-1}\sum_{t=1}^{n-1}(y_{t+1} - \hat{\mu}_n^* - \hat{\rho}_n^* y_t)^2$. In addition, $b = n(1-\rho)$ can be consistently estimated by $\hat{b}_n^* = n(1-\hat{\rho}_n^*)$, in view of Theorems 2.1 and

2.2. The performance of $\hat{b}_n^*$ is demonstrated via the empirical estimate,

$$\hat{\mathrm{E}}\left(\frac{\hat{b}_n^* - b}{b}\right) = \frac{1}{5000}\sum_{i=1}^{5000}\frac{\hat{b}_n^*(i) - b}{b},$$

of the relative bias $\mathrm{E}[(\hat{b}_n^* - b)/b]$, based on the data generated from 5000 simulation runs of model (1.1) with Beta(α,1) error, where $\rho \in \{0.86, 0.9, 0.95, 0.975, 0.99\}$, $\alpha \in \{1, 1.5, 2, 2.5, 3.5, 4\}$, and $\hat{b}_n^*(i)$ is $\hat{b}_n^*$ obtained in the ith simulation. Since our study is meant to be illustrative rather than exhaustive, we only focus on the sample size $n = 10000$. The results are summarized in Table 7. It is shown in Table 7 that all values of $\hat{\mathrm{E}}[(\hat{b}_n^* - b)/b]$ are quite close to 0, and $|\hat{\mathrm{E}}[(\hat{b}_n^* - b)/b]|$ is clearly smaller in the case of $\alpha < 2$ than in the case of $\alpha \geq 2$. This latter feature coincides with the fact that the convergence rate of $\hat{\rho}_n$ in the case of $\alpha < 2$ is faster than $\tilde{\rho}_n$.

Estimating c and α is much more involved than estimating b. While it seems feasible to perform kernel density estimation based on the AR residuals, $\hat{\varepsilon}_i = y_i - \hat{\rho}_n^* y_{i-1}$, to estimate c and α, the usual kernel estimators can be seriously biased when $0 < \alpha \leq 1$ because the corresponding density function is nonzero or even has a pole at the origin; see Marron and Ruppert (1994). Indeed, Marron and Ruppert (1994) suggested some sophisticated kernel estimation algorithms to reduce the boundary bias. However, consistency of the resulting estimators of c and α still seems difficult to establish when only (1.2) is assumed. In this connection, we also mention that a similar

Table 1: The values of $\hat{\mathrm{E}}[(\hat{b}_n^* - b)/b]$, with $n = 10000$, under model (1.1) with Beta$(\alpha, 1)$ errors.

	$\rho(b)$				
α	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002
1.5	-0.0003	-0.0003	-0.0003	-0.0003	-0.0003
2	-0.014	-0.013	-0.013	-0.013	-0.013
2.5	0.001	0.004	0.003	0.004	0.004
3	0.003	0.004	0.004	0.003	0.003
4	0.001	0.002	0.005	0.004	0.002

Table 2: The values of $\hat{\mathrm{E}}(\hat{\alpha}_n^*(m) - \alpha)$, with $n = 10000$, under model (1.1) with Beta$(\alpha, 1)$ errors.

		$\rho(b)$				
α	m	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	250	0.022	0.023	0.018	0.022	0.023
1.5	250	-0.013	-0.009	-0.006	-0.007	-0.009
2	500	-0.040	-0.040	-0.040	-0.040	-0.043
2.5	250	0.068	0.107	0.095	0.109	0.099
3	250	0.098	0.101	0.088	0.097	0.095
4	250	0.097	0.101	0.148	0.128	0.114

difficulty arises in constructing a confidence interval for ρ based on (1.5), in which α and c appear in the normalizing constant and α also appears in the limit. To bypass this difficulty, Datta and McCormick (1995) proposed an asymptotically pivotal quantity based on $\hat{\rho}_n$ and adopted a bootstrap procedure to consistently estimate the limit distribution of the proposed pivotal quantity.

Here, we take a somewhat nonstandard approach to estimate α and c. Note that (1.2) yields

$$\lim_{n\to\infty} P(n^{1/\alpha}\varepsilon_{(1)} > x) = \exp(-(c/\alpha)x^{\alpha}),$$

where $\varepsilon_{(j)}$ is the jth order statistic of $\{\varepsilon_1, \ldots, \varepsilon_n\}$, and hence $n^{1/\alpha}\varepsilon_{(1)}$ has the limiting Weibull density,

$$f_{(1)}(x) = \frac{\alpha}{\lambda^{\alpha}} x^{\alpha-1}\exp(-(x/\lambda)^{\alpha}), \tag{S2.31}$$

with shape parameter α and scale parameter $\lambda = (\alpha/c)^{1/\alpha}$. This motivates the following procedure for estimating α and c:

1. Produce the AR residuals: $\hat{\varepsilon}_{i+1} = y_{i+1} - \hat{\rho}_n^* y_i, i = 1, \ldots, n-1$.

2. Divide $\{1, \ldots, n\}$ into m subgroups, $\{1, \ldots, n_1\}, \ldots, \{n_{m-1}+1, \ldots, n_m\}$, where $n_i = \lfloor (n-1)/m \rfloor$ or $\lfloor (n-1)/m \rfloor + 1$ with $\lfloor a \rfloor$ denoting the largest integer $\leq a$.

3. Let $\hat{\varepsilon}_{(1)}(j)$ denote the smallest positive value among $\{\hat{\varepsilon}_{n_{j-1}+1}, \ldots, \hat{\varepsilon}_{n_j}\}$, $j = 1, \ldots, m$.

4. Use the Weibull density (S2.31) and $n_1^{1/\alpha}\hat{\varepsilon}_{(1)}(1), \ldots, n_m^{1/\alpha}\hat{\varepsilon}_{(1)}(m)$ to construct the maximum likelihood estimate $(\hat{\alpha}_n^*(m), \hat{\lambda}_n^*(m))$ of (α, λ).

5. Estimate c by $\hat{c}_n^*(m) = \hat{\alpha}_n^*(m)/(\hat{\lambda}_n^*(m))^{\hat{\alpha}_n^*(m)}$.

Under the stationary model (1.1), Hsiao, Huang, and Ing (2017) established the consistency of $(\hat{\alpha}_n(m), \hat{c}_n(m))$ regardless of whether $\alpha \leq 2$ or $\alpha > 2$, where $(\hat{\alpha}_n(m), \hat{c}_n(m))$ is $(\hat{\alpha}_n^*(m), \hat{c}_n^*(m))$ with $\hat{\varepsilon}_{i+1}$ replaced by the EV residual $y_{i+1} - \hat{\rho}_n y_i$, $m \to \infty$ and $n/m \to \infty$. This result enables them to asymptotically correctly identify the better estimator between $\hat{\rho}_n$ and $\tilde{\rho}_n$ in a data-driven fashion. The consistency of $(\hat{\alpha}_n^*(m), \hat{c}_n^*(m))$ under the near unit-root model (1.7) can also be established by an argument similar to that used in Hsiao, Huang and Ing (2017). The details, however, are not pursued here. In Table 8, the empirical estimate,

$$\hat{\mathrm{E}}(\hat{\alpha}_n^*(m) - \alpha) = \frac{1}{5000}\sum_{i=1}^{5000}(\hat{\alpha}_{n,i}^*(m) - \alpha),$$

of the bias of $\hat{\alpha}_n^*(m)$, $\mathrm{E}(\hat{\alpha}_n^*(m) - \alpha)$, is presented under the same scenarios as those in Table 7, where $\hat{\alpha}_{n,i}^*(m)$ is $\hat{\alpha}_n^*(m)$ obtained in the ith simulation. The tuning parameter m is set to 250 and 500 in our study. However, only the smaller one between $\hat{\mathrm{E}}(\hat{\alpha}_n^*(250) - \alpha)$ and $\hat{\mathrm{E}}(\hat{\alpha}_n^*(500) - \alpha)$ is reported in

Table 8. It remains for future research to choose m such that the resultant $\hat{\alpha}_n^*(m)$ has a better finite sample performance. With the same m as in Table 8, we present the empirical estimate, $\hat{\mathrm{E}}(\hat{c}_n^*(m) - c)$, of $\mathrm{E}(\hat{c}_n^*(m) - c)$ in Table 9. Table 8 reveals that $\hat{\alpha}_n^*(m)$ appears to be a reliable estimate of α because all values of $|\hat{\mathrm{E}}(\hat{\alpha}_n^*(m) - \alpha)|$ are small. On the other hand, we notice that $|\hat{\mathrm{E}}(\hat{\alpha}_n^*(m) - \alpha)|$ is larger in $\alpha > 1.5$ than $\alpha = 1.5$, which may be attributed to a slower convergence rate of $\hat{\rho}_n^*$ in the former case. In addition, perhaps due to a positive value of the density function at the origin, the performance of $\hat{\alpha}_n^*(m)$ in the case of $\alpha = 1$ also looks inferior to that in the case of $\alpha = 1.5$, although $\hat{\rho}_n^*$ in the former case has a faster convergence rate.

Table 9 shows that the performance of $\hat{c}_n^*(m)$ is in general unsatisfactory. In particular, all values of $\hat{\mathrm{E}}(\hat{c}_n^*(m) - c)$ are positive and are considerably larger than 0 for $\alpha > 2$. Taking a closer look at

$$\hat{c}_n^*(m) = \frac{\hat{\alpha}_n^*(m)}{m^{-1}\sum_{i=1}^m n_i[\hat{\varepsilon}_{(1)}(i)]^{\hat{\alpha}_n^*(m)}}, \tag{S2.32}$$

we found that a non-negligible portion of $\{n_i[\hat{\varepsilon}_{(1)}(i)]^{\hat{\alpha}_n^*(m)}\}$ concentrates near 0. As a result, the denominator on the right-hand side of (S2.32) tends to underestimate $\lambda^\alpha = \alpha/c$, and hence $\hat{c}_n^*(m)$ tends to overestimate c, as observed in Table 9. To remedy this difficulty, we suggest an alternative, $\hat{c}_n^*(m,r)$, which is $\hat{c}_n^*(m)$ with the denominator replaced by the sample mean

Table 3: The values of $\hat{\mathrm{E}}(\hat{c}_n^*(m) - c)$, with $n = 10000$ and the same m as those in Table 8, under model (1.1) with Beta$(\alpha, 1)$ errors.

	$\rho(b)$				
α	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	0.152	0.152	0.151	0.160	0.146
1.5	0.033	0.059	0.049	0.071	0.055
2	0.097	0.109	0.111	0.105	0.077
2.5	0.484	0.596	0.698	0.703	0.539
3	0.517	0.451	0.388	0.473	0.455
4	0.591	0.578	0.565	0.560	0.551

from the highest $(1-r)\%$ of the elements of $\{n_i[\hat{\varepsilon}_{(1)}(i)]^{\hat{\alpha}_n^*(m)}, i = 1, \dots m\}$. Under the same simulation setting as Table 9, we compute the empirical estimate, $\hat{\mathrm{E}}(\hat{c}_n^*(m,r) - c)$, of $\mathrm{E}(\hat{c}_n^*(m,r) - c)$, and report the smallest one among $\hat{\mathrm{E}}(\hat{c}_n^*(m,r) - c), r = 5, 10, 15, 20, 25, 30$; see Table 10. Table 10 shows that all values of $|\mathrm{E}(\hat{c}_n^*(m,r) - c)|$ are not distant from 0, and clearly smaller than $|\hat{\mathrm{E}}(\hat{c}_n^*(m) - c)|$.

We now return to the more practical situation where $\alpha > 2$ or $\alpha \leq 2$ is unknown. In this case, we suggest the following rule:

Rule III.

1. Judge $\alpha > 2$ if $\hat{\alpha}_n(m) - \xi > 2$,

Table 4: The values of $\hat{\mathrm{E}}(\hat{c}_n^*(m,r)-c)$, with $n=10000$ and the same m as those in Table 8, under model (1.1) with Beta$(\alpha,1)$ errors.

		$\rho(b)$				
α	r	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	25	0.014	0.011	-0.006	0.006	0.014
1.5	5	-0.005	-0.013	-0.031	0.015	0.013
2	5	0.023	0.018	0.021	0.015	0.005
2.5	20	-0.023	0.082	0.093	0.133	0.062
3	15	0.077	0.069	0.021	0.066	0.060
4	15	-0.046	-0.023	0.117	0.050	0.088

Table 5: The values of F, with $n=10000$, under model (1.1) with Beta$(\alpha,1)$ errors.

	$\rho(b)$				
α	0.86(1400)	0.9(1000)	0.95(500)	0.975(250)	0.99(140)
1	1.000	1.000	1.000	1.000	1.000
1.5	1.000	1.000	1.000	1.000	1.000
2	0.978	0.984	0.980	0.982	0.986
2.5	0.922	0.930	0.938	0.926	0.908
3	1.000	0.996	0.998	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000

2. Judge $\alpha \leq 2$ if $\hat{\alpha}_n(m) - \xi \leq 2$,

where $\hat{\alpha}_n(m)$ is defined previously and ξ is a prescribed small positive number. In Table 11, with the same scenarios as those in Table 7, we report the percentage, F, of Rule III (with $m = 500$ and $\xi = 0.14$) making correct judgements, where

$$F = \sharp(\{i : 1 \leq i \leq 5000, I(\hat{\alpha}_{n,i}(500) - 0.14 > 2) = I(\alpha > 2)\})/5000,$$

$n = 10000$ and $\hat{\alpha}_{n,i}(500)$ denotes $\hat{\alpha}_n(500)$ obtained in the ith simulation. Note that $\xi = 0.14$ is an approximation of $2\sigma_{\mathrm{MLE}}/\sqrt{500}$, where σ^2_{MLE} is the limiting variance of the MLE of α of the Weibull density (A.6) calculated at $\alpha = 2$ and $\lambda = 1$. As shown in Table 11, all values of F are near 1, in particular when $\alpha < 2$ or $\alpha > 2.5$. This result implies that Rule III provides a reliable decision about whether or not $\alpha > 2$, thereby allowing one to carry out the aforementioned estimates of α and c in practice.

Finally, we want to reiterate that this section is exploratory in nature, and there remain a number of unsettled issues (e.g., the choices of m and r in $\hat{\alpha}^*_n(m)$ and $\hat{c}^*_n(m, r)$) worthy of further investigation. On the other hand, our simulation study suggests that the notoriously difficult problem of estimating α and c in the distribution of ε_t can be somewhat alleviated through the proposed estimates, $\hat{\alpha}^*_n(m)$ and $\hat{c}^*_n(m, r)$, provided m and r are properly given.